Graphics, Vision and Image Processing Journal, ISSN 1687-398X, Volume 17, Issue 2, ICGST LLC, Delaware, USA, Nov. 2017

www.icgst.com

A Precise Technique for Hand Gesture Recognition

Ayush Purohit[1], Shardul Singh Chauhan[2]
Software Engineer, Larsen and Toubro Infotech, Mumbai, India[1]
Assistant Professor, Department of CSE, Northern India Engineering College, New Delhi, India[2]
ayushpurohit@outlook.com, shardulchauhan007@gmail.com

Abstract
Vision based methodologies provides a more natural and proficient result when contrasted with traditional strategies which have been utilized for hand gesture recognition. In this paper, we proposed a video based hand gesture recognition. Our approach commences by acquiring the video frame from a source and converting it into 2D binary frame using YCbCr color space. We implemented opening and closing operations to filter the noise from the frame. In order to track and segment the hand gesture we used Kalman filter and convex hull along with convexity defects for detecting hand regions from the frame. Our framework can perceive six kinds of hand gestures at present time.

Keywords: *Computer Vision, Convex Hull, Convexity Defect, Kalman Filter.*

Nomenclature

SVM	Support Vector Machine
CNN	Convolutional Neural Network
SAD	Sum of Absolute Differences
HOG	Histogram of oriented Gradients
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis

1. Introduction
Gesture recognition is a process of deciphering and comprehending the human gestures by implementing various algorithms. Gesture Recognition has been an area where a colossal measure of examination has been done which has numerous applications. An assortment of methodologies has been proposed for the procedure of gesture recognition. Data glove based methodology makes the utilization of sensor gadgets for digitalization of both hand and additionally finger movements into multi parametric information. Movement construct hand division approaches depend with respect to the supposition that the elements vital for gestures will be connected with gestures. Vision-based methodologies share the issue identified with the caprices of low-level division. Most of the image processing techniques are in light of two fundamental techniques: machine learning and rules.

A vision based hand gesture recognition system is proposed in [1] which uses scale space highlight discovery. In this work the first step is to make use of a specific hand gesture in order to detect the hands followed by tracking. The segmentation of hands is done using color cues and motion. Finally a scale space feature detection technique is used for integration in recognition of gestures. Jesus et al, in [2] examines depth based hand gesture recognition. The point has been to highlight the gesture classification strategies and additionally hand restriction techniques. Here a detailed study of 37 papers have been made for comparing various depth based gesture recognition systems on the basis of various aspects like hand localization, the effects of low cost Kinect, OpenNI software libraries and so on. A video based hand gesture recognition method has been implemented in [3]. The work focuses on recognition of hand gestures on a video stream. The proposed system focuses on two procedures namely the hand gesture detection and hand gesture recognition. The hand detection begins by locating the hands in the video frames with the help of blue rectangles by implementing Viola Jones technique. The hand gesture recognition begins with the Hu invariant moments feature vectors which are extracted from the detection of hand gestures and then trained and classified using SVM.

Another methodology is proposed in [4] utilizes modified census transform to highlight extraction process for gesture recognition. The claim to fame of the transform is that it is enlightenment invariant. Finally, a direct classifier is used for recognizing hand gestures. A video based hand gesture recognition technique is suggested in [5]. Initially a user hand gesture video is captured and stored in the hard disk. The videos captured are read by the system one by one and converted in the form of binary images. Then a 3D Euclidian space is created of the binary values obtained. For the training a feed forward neural network training method and for classification back propagation neural network is used. In [6], gesture recognition method is proposed which uses feed forward neural networks alongside back propagation for classifying the extracted features. The work compares various hand gesture recognition techniques by making the use of MATLAB. The use and implementation of skin detection and edge detection algorithms are also studied. Reference in [7], concentrates on the utilization of CNN for hand gesture recognition by making use of images captured by camera. To make the system robust, calibration of hand position, orientation and skin model are applied for obtaining the training as well as testing data for CNN. The Gaussian mixture model algorithm is used for training of the skin model. The calibrated images so obtained are used for the purpose of training the CNN.

Xianghua Li proposed thinning method which involves SAD to compute matching regions [8]. A depth map is implemented in the portion of hand detection that makes the use of sum of absolute differences technique for detection of the object located in foreground. The frame is converted into YCbCr space and then convex hull is computed to extract region of interest. The background image in the obtained region of interest is removed so that the foreground image can

be received that is hand image. A blob labeling method algorithm is used for obtaining the clear hand image. The feature point extracted using thinning algorithm is used to recognize hand gesture. Similar approach is used by Amiraj in [9], uses convex hull and convexity defects to count the number of fingers in video. The primary step is to capture the video and use it as an input for the system. The video is converted into frames and thresholding is applied to separate the hands from the background. Contours are used to find out the location of hands in the video frames. The algorithms like convex hull and convexity defects are implemented for detection and extraction of hands from the input. Then by making the use of various rules the hand gestures are classified. In [10], proposed automated method to recognize hand gestures in varying backgrounds. Skin color detection method has been used to figure out the hand region from the complex background. A series of morphological operations are implemented to extract the contour which is used to recognize finger tips. The angle of the fingertips is used for marking the fingertips. The technique shows the accuracy of the system with low computation cost. Yafei used HOG transform to extract hand features which are then reduced to 9D subspace using PCA-LDA [11]. The hand regions are finding out by combining an adaptive skin color detection algorithm along with the motion detection. The distance between the features of projections and each class of gesture is calculated. The extracted features are then classified using nearest neighbor to identify the gesture. The use of hands instead of mouse as an input appears to be an instinctive choice for man machine interaction.

In this paper, we used convex hull and convexity defects to describe the hand gestures. The hands are firstly detected using skin color and various morphological operations are used to extract the features using convex hull properties. For the purpose of tracking, Kalman filter is used to track the location of hands in the video frames. The classification is done on the basis of the specified rule set. Finally the results of the proposed technique have been tabulated which indicates the precision of the system.

2. Hand Detection

In order to locate the hand gesture in a video frame efficiently, skin color detection and region of interest are computed.

Skin Color Detection

Skin color detection is a procedure of identifying the region of interests within the spectrum of skin colored pixels in an image or a video frame. This methodology is utilized in various approaches which incorporate distinguishing a face, object, hand, etc. in diversified expanses.

Due to vacillating background conditions & luminance components, we erected our skin color model in YCbCr color space in order to approximate the chromaticity of skin. This computation involves conversion of RGB to YCbCr color space and eliminating the luminance component to compose the skin color more robust to illumination. The histogram of the resulting 2D color vector has produced the region of interest which shows a strong peak at the skin color. This conversion step is explained using a diagram as represented in figure 1.

The YCbCr conversion of a given pixel from RGB can be deduced by the following matrix I:

$$\begin{bmatrix} Y \\ Cr \\ Cb \end{bmatrix} = [\text{R G B}] \begin{bmatrix} 0.299 & -0.1689 & 0.4998 \\ 0.587 & -0.3317 & 0.4185 \\ 0.114 & -0.5006 & -0.0813 \end{bmatrix} \qquad \text{(I)}$$

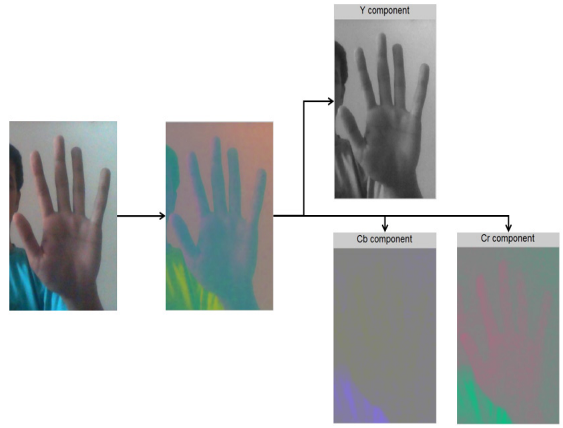


Figure 1: Step1: Conversion of RGB to YCbCr color space. Step 2: Separating Y, Cb and Cr components from YCbCr frame.

3. Motion Detection

To the resultant 2D gray scale color vector, morphological transformations are performed. We initiated the process by thresholding the grayscale framework. This method reorganizes a grayscale image to a bi-level image and extracts the pixels representing the hand or an object. A median filter with a kernel 15 x 15 is used to filter the noise from the resulted frame. A combination of morphological operations which consist of binary opening and closing, are applied over the image to suppress the remaining noise using a square kernel.

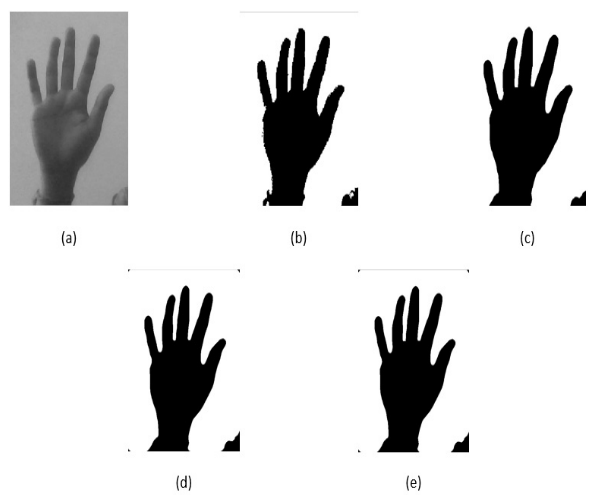


Figure 2: a) Grayscale Image. b) Threshold Operation. c) Median Filter Operation. d) Opening Operation. e) Closing Operation.

The opening of image I by kernel H can be computed as:

$$(I o \text{H}) = (I \oplus H) \Theta H \qquad (1)$$

To the resulting frame, we performed thresholding operation to acquire an optimal frame for computing hand gesture features. A series of morphological operations implemented over the video frame is shown in figure 2.

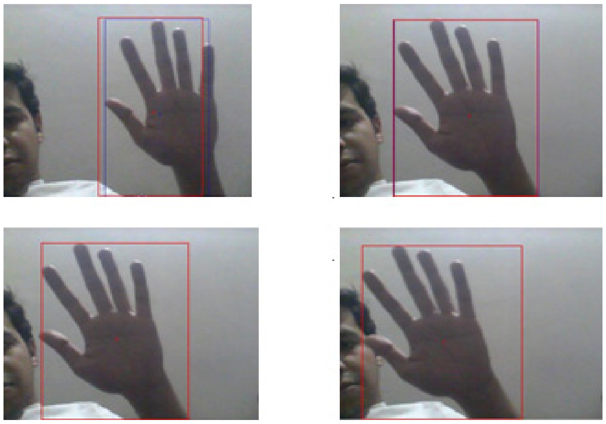


Figure 3: Sequence of frames extracted from video using Kalman Filter.

Hand Tracking

To track hand gestures in real time, we implemented an optimal estimation framework catered by Kalman filter which is extensively adopted for tracking objects because of its small computational requirements, elegant recursive properties, uncertainty analysis and prognosis of subsequent frames [12] [13]. In this paper, Kalman filter is employed to predict the location of hand gesture in a frame. The Kalman filter follows a two-step procedure for hand tracking, that are control and measurement update. The control update can be used for estimation of the state with the previous state and vector, while the measurement update is used for correcting the sensor information based upon the state. To finally predict the position of hand in the frame, we blend the Gaussian results produced from prediction and measurement to obtain the position of the hand as shown in figure 3.

4. Projection into Palm Plane

To project into palm plane, contours are configured around the black bead of the hand developed after segmenting the frame. It is possible that system might detect multiple contours which are produced due to noise in the background. An assumption is made that contours produced by the noise are smaller in size compared to contour of the hand. Therefore, we scrutinized the biggest contour in the frame which is used for further processing. This method thus removes the possibility of considering any contour formed due to noise.

Convexity Detection

The final approach of our system is to detect convexity points from the extracted contour. This methodology endeavors to detect convex hull and convexity points from the contour. The convex hull illustrates the extrinsic contour of the hand such that all the contour specks are within the convex hull.

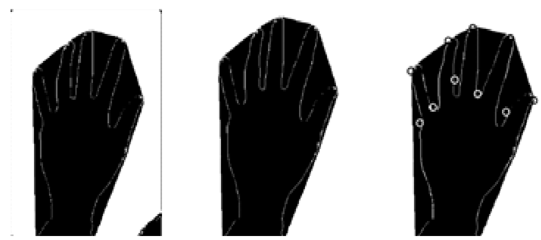
To extract the convex hull, we approximated the hand contour with a minimum parameter polygon resulting in dwindling of undesirable convexity specks. We used Douglas-Peucker algorithm for smoothing the boundary which recursively joins first and last vertices of the polygonal line segment by finding the vertex furthest from it.

To estimate convex hull points of the approximation polygon, we implemented a simple and intuitive Sklansky's algorithm. This graph based algorithm is based on stack, which in the extreme includes the vertices of the convex hull. It considers three vertices: top stack vertex, new vertex, second to top of the stack vertex. The top stack vertex is rebuffed if trio forms a right turn.

Convexity defects are computed by measuring distance between the farthest point and convex hull. The resulting frame is filtered by rejecting the convex points which are not present near finger tips. This is done by computing the centroid of enclosed polygon. If any convex point whose height is less, then height of the center of the palm was filtered out.

Hand Gesture Recognition

This application is developed to identify the number of fingers operating in a hand gesture. To classify the number of fingers distinguishable in the frame, we used feature extracted from frames and counted the number of convex and convexity defect points. Figure 4 indicates the use of convex hull and convexity defects to find out the hand points that are needed for recognizing hand gestures.



(a) (b) (c)

Figure 4: (a) Convex Hull of the frame. (b) Extracted Contour. (c) Convex and Convexity Defect Points.

Finger Counting

Using polylines drawn around the hand, we computed the approximate centroid of the hand. For any of the parameter to correctly satisfy the prerequisite, the 'x' number of convex hull points should lay outside a threshold range from the centroid of the hand. In order to recognize the number of fingers, one of the following parameters should be satisfied as shown in Table 1:

No. of fingers	Convex Hull Points (x)	Convexity Defect Points (y)
0	Exactly 0	Exactly 0
1	Exactly 1	Exactly 0
2	Exactly 2	At least 1
3	Exactly 3	At least 1
4	Exactly 4	At least 1
5	Exactly 5	At least 1

Table 1: Condition for recognizing finger counts

5. Experimental Results

In this model, there are certain constraints which need to be satisfied for recognizing the hand gesture and count the number of fingers. The system also maintains the tracks of the hand gesture which uses Kalman filter. In figure (5) and (6), shows the current working model which can trail the hand and recognizes limited number of finger counts. In order to find out the classification rate of the system a set of 20 videos are used for each hand gesture. The aim was to ensure that the arrangement of videos contain enough data with a specific end goal to depict a specific hand gesture.

The set involves videos which delineates a solitary hand performing gestures where hand ought to possess the significant locale. Table 2 indicates the classification results of the system.

		Result of Classification						Unrecognized	Error Percentage
Inputs	Class of Gestures	0	1	2	3	4	5		
	0	19	0	0	0	0	0	1	5
	1	0	20	0	0	0	0	0	0
	2	0	0	20	0	0	0	0	0
	3	0	0	0	20	0	0	0	0
	4	0	0	0	0	20	0	0	0
	5	0	0	0	0	0	20	0	0

Table 2: Classification Results

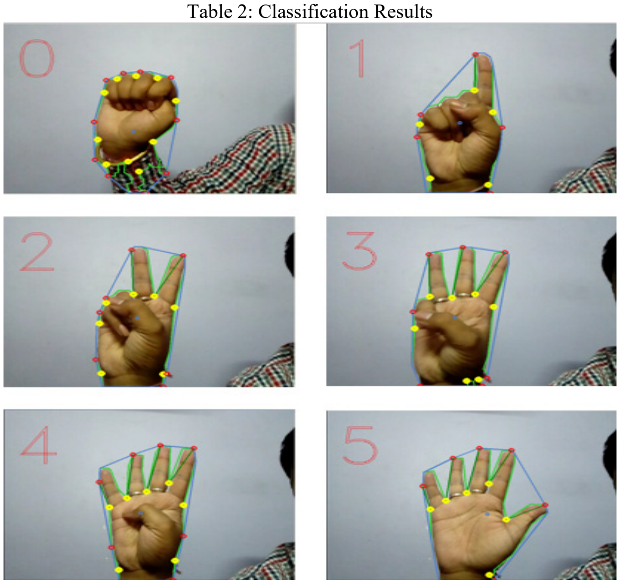


Figure 5: Finger Counting using Convex Hull & Convexity Defects.

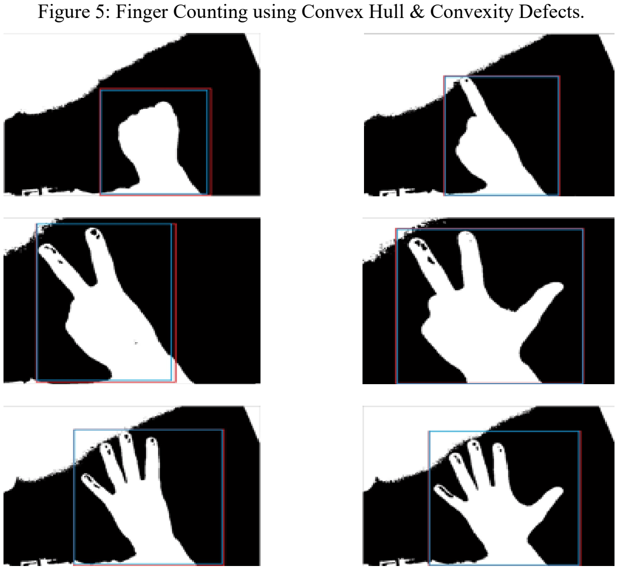


Figure 6: Hand Tracking in Binary Video Using Kalman Filter.

6. Conclusion and Future Work

In this paper, we presented a vision-based hand gesture recognition system which operates on real time videos on an average PC using low cost cameras. The proposed method is currently used to count limited number of fingers with a high classification rate under various constraints. The future work involves recognizing multiple hands in a given frame, a rotation and orientation independent gesture recognition and a more efficient and flexible man-machine interaction which can be used in real life applications.

References

[1] Yikai Fang, Kongqiao Wang, Jian Cheng, Hanqing Lu, A Real-Time Hand Gesture Recognition Method, *Proc. IEEE International Conference on Multimedia and Expo*, 2007, 995-998.

[2] Jesus Suarez, Robin R. Murphy, Hand Gesture Recognition with Depth Images: A Review, *Proc. IEEE Ro-Man: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, 2012, 411-417.

[3] Liu Yun, Zhang Peng, An Automatic Hand Gesture Recognition System Based on Viola-Jones Method and SVM's, *2nd IEEE International Workshop on Computer Science and Engineering*, 2009, 72-76.

[4] Agnes Just, Yann Rodriguez, Sebastien Marcel, Hand Posture Classification and Recognition using the Modified Census Transform, *Proceedings of the 7th IEEE International Conference on Automatic Face and Gesture Recognition*, 2006, 351-356.

[5] Murthy, G.R.S, Jadon, R.S, Hand gesture recognition using neural networks, *Proc. IEEE 2nd International Advance Computing Conference*, 2010, 134-138.

[6] Yewale, S.K, Bharne, P.K, Hand gesture recognition using different algorithms based on artificial neural network, *IEEE International Conference on Emerging Trends in Networks and Computer Communications*, 2011, 287-292.

[7] Hsien-I Lin, Ming-Hsiang Hsu, Wei-Kai Chen, Human hand gesture recognition using a convolution neural network, *IEEE International Conference on Automation Science and Engineering*, 2014, 1038-1043.

[8] Xianghua Li, Jun-ho An, Jin-hong Min and Kwang-Seok Hong, Hand gesture recognition by stereo camera using the thinning method, *IEEE International Conference on Multimedia Technology (ICMT)*, 2011, 3077-3080.

[9] Amiraj Dhawan, Vipul Honrao, *Implementation of Hand Detection based Techniques for Human Computer Interaction, International Journal of Computer Applications, 18(6)*, 2013, 0975 – 8887.

[10] Xiaotang Wen, Yanyu Niu, A method for hand gesture recognition based on morphology and Fingertip-Angle, *IEEE 2nd International Conference on Computer and Automation Engineering (ICCAE)*, 2010, 688-691.

[11] Yafei Zhao, Weidong Wang, Yuehai Wang, A real time hand gesture recognition method, *IEEE International Conference on Multimedia and Expo*, 2007, 995-998.

[12] C. K. Chui, G. Chen, *Kalman Filtering with real-time applications* (Berlin, Springer-Verlag).

[13] Shiuh-Ku Weng, Chung-Ming Kuo, Shu-Kang Tu, Video object tracking using adaptive Kalman filter, Journal of Visual Communication and Image Representation, volume 17, issue 6, Dec 2006.

Biographies

Mr. Ayush Purohit is currently working as a Software Engineer in Larsen and Toubro Infotech, Mumbai. He completed his BTech in Computer Science and Engineering and MTech in Artificial Intelligence and Artificial Neural Networks from University of Petroleum and Energy Studies, Dehradun. His area of interest includes Computer Vision, Data Visualization and Business Process Automation.

Mr. Shardul Singh Chauhan is currently working as an Assistant Professor in Northern India Engineering College, New Delhi. He completed his BTech in Computer Science and Engineering and MTech in Artificial Intelligence and Artificial Neural Networks from University of Petroleum and Energy Studies, Dehradun. His area of interest includes Digital Image Processing, Artificial Neural Networks, Expert Systems and Pattern Recognition.